correctly linked cysteine bridges; 9) concentration of the proinsulin solution; 10) chromatographic purification of the concentrated proinsulin solution; 11) enzymatic cleavage of the proinsulin in order to obtain human insulin; and 12) chromatographic purification of the resulting human insulin (EP 0,055,945). Disadvantages of the this process are the numerous procedural steps and the losses in the purification steps. which lead to a low yield of insulin. From the step of the isolated fusion protein via cyanogen bromide cleavage, sulfitolysis and purification of the proinsulin, a loss of proinsulin of up to 40% is to be expected (EP 0,055,945). On the other hand, the yield of recombinantly producing insulin, or its derivatives, can be significantly increased if the number of the necessary procedural steps are significantly reduced.

One objective of the present invention was to develop a recombinant process for obtaining human insulin with correctly linked cysteine bridges with fewer necessary procedural steps, and hence resulting higher yield of human insulin. Another objective of the present invention was to develop an insulin-precursor-containing chimeric protein that can be used in the above process. Still another objective of the present invention was to develop an assay for screening an amino acid sequence, when linked to an insulin precursor via peptidyl bond, will improve folding of the insulin precursor.

Applicants have searched for peptide sequences that would not only protect insulin sequences from the intracellular degradation by microorganism host, but also, compared to the then existing human insulin expression system, possess the following advantages: when linked to an insulin precursor via peptidyl bond, 1) promotes the folding of the fused insulin precursor; 2) facilitates the solubility of the fusion protein and decrease the intermolecular interactions among the fusion proteins, thus allowing folding of the fused insulin precursor at a commercially significant high concentration; 3) eliminates the procedural steps of cyanogen bromide cleavage, oxidative sulfitolysis and the related purification steps; and 4) eliminates the use of high concentration of mercaptan or the use of hydrophobic absorbent resins.

Applicant found, surprisingly, that linking an IMC like sequence to an insulin precursor via one or more cleavable amino acid residues accomplish the objectives of the present invention. The IMC like sequence has certain characteristics of an IMC sequence such as helping the target protein folding, containing higher percentage of charged amino acid residues than its target protein, having polarized distribution of the charged amino acid residues and having a sequence that appears to be "tailor-made" for the target protein. However, the IMC like sequence used in present invention is different from an IMC sequence in several key aspects. First, the IMC like sequence is heterogeneous to the target protein, i.e., not a propeptide of the target protein. For example, if an insulin

precursor is a target protein to be folded, an IMC like sequence is not the insulin precursor or a portion thereof. In addition, the size of the IMC like sequence is from about 20 to about 200 amino acid residues.

Additionally, contrary to the teaching in the prior art (Castellanos-Serra et al., *FEBS Letters*, 1996, 378:171-176), Applicant found, surprisingly, that including, within the IMC like sequence, one or more cleavable amino acid residues which are identical to the one or more cleavable amino acid residues that separate the IMC like sequence and an insulin precursor allows fragmented removal of the IMC like sequence after folding, hence, simplifying down-stream purification steps.

For clarity of disclosure, and not by way of limitation, the detailed description of the invention is divided into the subsections which follow.

4.1. NUCLEIC ACIDS ENCODING THE CHIMERIC PROTEIN DISCLOSED IN SECTION 4.2.

The present invention provides an isolated nucleic acid comprising a nucleotide sequence encoding the chimeric protein disclosed in Section 4.2.

In a specific embodiment, the present invention provides an isolated nucleic acid comprising a nucleotide sequence encoding the chimeric protein having the amino acid sequence of SEQ ID NO:6.

In another specific embodiment, the present invention provides an isolated nucleic acid comprising a nucleotide sequence encoding the chimeric protein having the amino acid sequence of SEQ ID NO:7.

In a preferred embodiment, the present invention provides an isolated DNA molecule comprising a nucleotide sequence encoding the chimeric protein disclosed in Section 4.2.

In another preferred embodiment, the present invention provides an isolated nucleic acid comprising a nucleotide sequence complementary to the nucleotide sequence encoding the chimeric protein disclosed in Section 4.2.

In still another specific embodiment, the present invention provides an isolated nucleic acid hybridizable to the nucleotide sequence encoding the first, second and third peptidyl fragments of the DNA encoding the chimeric protein disclosed in Section 4.2.

The nucleic acid comprising a nucleotide sequence encoding the chimeric protein disclosed in Section 4.2., or any fragments, analogues or derivatives thereof, can be obtained by any method(s) known in the art. The nucleic acid may be chemically synthesized entirely. Alternatively, the nucleic acid encoding each fragment of the chimeric protein, i.e., the first, second or third peptidyl fragment, may be obtained by

molecular cloning or may be purified from the desired cells. The nucleic acid encoding each fragment of the chimeric protein may then be chemically or enzymatically ligated together to form the nucleic acid comprising a nucleotide sequence encoding the chimeric protein disclosed in Section 4.2., or any fragments, analogues or derivatives thereof.

Any human cell potentially can serve as the nucleic acid source for the isolation of *hGH* nucleic acids. Any mammalian cell potentially can serve as the nucleic acid source for the isolation of *insulin* nucleic acids. The nucleic acid sequences encoding insulin can be isolated from mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional rodent or primate sources, etc.

The DNA may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell (*See*, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will contain only exon sequences. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from cDNA, cDNA is generated from totally cellular RNA or mRNA by methods that are well known in the art. The gene may also be obtained from genomic DNA, where DNA fragments are generated (e.g. using restriction enzymes or by mechanical shearing), some of which will encode the desired gene. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once a nucleic acid comprising a nucleotide sequence encoding the chimeric protein disclosed in Section 4.2., or any fragments, analogues or derivatives thereof, has been obtained, its identity can be confirmed by nucleic acid sequencing (by any method well known in the art) and comparison to the known sequences. DNA sequence analysis can be performed by any techniques known in the art, including but not limited to the method of Maxam and Gilbert (Maxam and Gilbert, 1980, *Meth. Enzymol.*, 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, *Proc. Natl. Acad. Sci. U.S.A.*, 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), use of an automated DNA sequenator (*e.g.*, Applied Biosystems, Foster City, CA) or the method described in PCT Publication WO 97/15690.